

# Literaturempfehlungen für den «CAS Translation Technology and AI»

Johannes Graën

Oktober 2020

## Zur Satzalignierung

- historisch, erste Alignierungsexperiment und -algorithmen: (Brown, Lai u. a. 1991; Gale und Church 1991)
- komplexere Modellierung: (Simard und Plamondon 1998; Moore 2002)
- Publikation zu `hunalign`: (Varga u. a. 2005)
- Publikation zu `bleualign`: (Sennrich und Volk 2010)
- Alignierungsalgorithmus für asymmetrische parallele Korpora: (Braune und Fraser 2010)
- Alignierung mithilfe von Information-Retrieval-Techniken: (Costa-jussà und Banchs 2011)
- Evaluierung/Vergleich verschiedener Alignierungstools: (Abdul-Rauf u. a. 2012; Torres-Ramos und Garay-Quezada 2015)

## Zur Wortalignierung

- die Einführung der wegweisenden «IBM-Modelle»: (Brown, Della Pietra u. a. 1993)
- Verbesserung mittels HMM-Modellen: (Vogel u. a. 1996)
- Alignierungstool «GIZA++»: (Al-Onaizan u. a. 1999)
- Wortalignierung mittels Triangulation: (Borin 2000)
- Vergleich von Wortalignierungsmodellen: (Och und Ney 2003)
- Identifikation von Mehrwortausdrücken mithilfe von Wortalignierung: (Zarrieß und Kuhn 2009)
- Typologischer Sprachvergleich mittels Wortalignierung: (Mayer und Cysouw 2012)

## Zu ‹Multilingual Search Tools›

- Vergleich verschiedener Werkzeuge zur Suche in parallelen Korpora: (Volk u. a. 2014)
- Unser Tool für die Exploration von Alignierungen in multiparallelen Korpora: (Clematide u. a. 2016; Graën, Sandoz u. a. 2017)
- das Parallelkorpora und das Suchwerkzeug PaGeS: (Doval Reixa 2016, 2018)
- Suche von Übersetzungswahrscheinlichkeiten mit Korpusbeispielen: (Graën und Schneider 2020)

## Andere Quellen zum Thema Alignierung

- Buch zu Alignierung paralleler Korpora: (Tiedemann 2011)
- OPUS – eine grosse Sammlung paralleler Korpora: (Tiedemann 2012)
- Modellierung von Alignierung mittels bayesscher Statistik: (Östling 2015)
- Alignierungsalgorithmen und verschiedene Werkzeuge, die Alignierung für theoretische und praktische Fragestellungen einsetzen, in meiner Dissertation: (Graën 2018)

## Literatur

- Abdul-Rauf, S., M. Fishel, P. Lambert, S. Noubours und R. Sennrich (2012). “Extrinsic Evaluation of Sentence Alignment Systems”. In: *Proceedings of the Workshop on Creating Cross-language Resources for Disconnected Languages and Styles (CREDISLAS)* (Istanbul), S. 6–10.
- Borin, L. (2000). “You’ll Take the High Road and I’ll Take the Low Road: Using a Third Language to Improve Bilingual Word Alignment”. In: *Proceedings of the 18th International Conference on Computational Linguistics (COLING)*. Saarbrücken, S. 97–103.
- Braune, F. und A. Fraser (2010). “Improved Unsupervised Sentence Alignment for Symmetrical and Asymmetrical Parallel Corpora”. In: *Proceedings of the 23rd International Conference on Computational Linguistics (COLING): Posters*. Association for Computational Linguistics (ACL), S. 81–89.
- Brown, P. F., V. J. Della Pietra, S. A. Della Pietra und R. L. Mercer (1993). “The Mathematics of Statistical Machine Translation: Parameter Estimation”. In: *Special issue on using large corpora: II* 19.2, S. 263–311.
- Brown, P. F., J. C. Lai und R. L. Mercer (1991). “Aligning sentences in parallel corpora”. In: *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL)*, S. 169–176.

- Clematide, S., J. Graën und M. Volk (2016). “Multilingwis – A Multilingual Search Tool for Multi-Word Units in Multiparallel Corpora”. In: *Computerised and Corpus-based Approaches to Phraseology: Monolingual and Multilingual Perspectives – Fraseologia computacional y basada en corpus: perspectivas monolingües y multilingües*. Hrsg. von G. C. Pastor. Geneva: Tradulex, S. 447–455.
- Costa-jussà, M. R. und R. E. Banchs (2011). *Sentence alignment by means of cross-language information retrieval*. INTECH Open Access Publisher.
- Doval Reixa, I. (2016). “Bilingual Parallel Corpora for Linguistic Research”. In: *CILC2016. 8th International Conference on Corpus Linguistics*. Hrsg. von A. M. Ortiz und C. Pérez-Hernández. Bd. 1. EPiC Series in Language and Linguistics, S. 88–96.
- (Jan. 2018). “Das PaGeS-Korpus, ein Parallelkorpus der deutschen und spanischen GegenwartsspracheEl corpus PaGeS”. In: *Revista de Filología Alemana* 26, S. 181–197.
- Gale, W. A. und K. W. Church (1991). “A Program for Aligning Sentences in Bilingual Corpora”. In: *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL)* (Berkeley, California). Stroudsburg, PA, USA, S. 177–184.
- Graën, J. (2018). “Exploiting Alignment in Multiparallel Corpora for Applications in Linguistics and Language Learning”. Diss. University of Zurich. Im Druck.
- Graën, J., D. Sandoz und M. Volk (2017). “Multilingwis2 – Explore Your Parallel Corpus”. In: *Proceedings of the 21st Nordic Conference of Computational Linguistics (NODALIDA)*. Linköping Electronic Conference Proceedings 131. Linköping University Electronic Press, Linköpings universitet, S. 247–250.
- Graën, J. und G. Schneider (2020). “Exploiting Multiparallel Corpora as a Measure for Semantic Relatedness to Support Language Learners”. In: *Strategies and Analyses of Language and Communication in Multilingual and International Contexts*. Hrsg. von D. Levey. Cambridge Scholars Publishing, S. 153–167.
- Mayer, T. und M. Cysouw (2012). “Language comparison through sparse multilingual word alignment”. In: *Proceedings of the Joint Workshop of Visualization of Linguistic Patterns (LINGVIS) & Uncovering Language History from Multilingual Resources (UNCLH)*. Association for Computational Linguistics (ACL), S. 54–62.
- Moore, R. C. (2002). “Fast and accurate sentence alignment of bilingual corpora”. In: *Proceedings of the 5th Conference of the Association for Machine Translation in the Americas (AMTA)*. Springer, S. 135–144.
- Och, F. J. und H. Ney (2003). “A Systematic Comparison of Various Statistical Alignment Models”. In: *Computational Linguistics* 29.1, S. 19–51.
- Al-Onaizan, Y. u. a. (1999). *Statistical Machine Translation*. Techn. Ber. Johns Hopkins University Summer Workshop.
- Östling, R. (2015). “Bayesian Models for Multilingual Word Alignment”. Diss. Stockholm University.

- Sennrich, R. und M. Volk (2010). “MT-based sentence alignment for OCR-generated parallel texts”. In: *Proceedings of the 9th Conference of the Association for Machine Translation in the Americas (AMTA)*.
- Simard, M. und P. Plamondon (1998). “Bilingual Sentence Alignment: Balancing Robustness and Accuracy”. In: *Machine Translation 13.1*, S. 59–80.
- Tiedemann, J. (2011). *Bidtext Alignment*. Bd. 4. Synthesis Lectures on Human Language Technologies 2. Morgan & Claypool.
- (2012). “Parallel Data, Tools and Interfaces in OPUS”. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC)* (Istanbul).
- Torres-Ramos, S. und R. E. Garay-Quezada (2015). “A Survey on Statistical-based Parallel Corpus Alignment”. In: *Research in Computing Science* 90, S. 57–76.
- Varga, D., L. Németh, P. Halácsy, A. Kornai, V. Trón und V. Nagy (2005). “Parallel corpora for medium density languages”. In: *Proceedings of Recent Advances in Natural Language Processing (RANLP)* (Borovets), S. 590–596.
- Vogel, S., H. Ney und C. Tillmann (1996). “HMM-based Word Alignment in Statistical Translation”. In: *Proceedings of the 16th Conference on Computational Linguistics*. Bd. 2. Association for Computational Linguistics (ACL), S. 836–841.
- Volk, M., J. Graën und E. Callegaro (2014). “Innovations in Parallel Corpus Search Tools”. Englisch. In: *Proceedings of the 9th International Conference on Language Resources and Evaluation (LREC)* (Reykjavik). Hrsg. von N. Calzolari u. a. European Language Resources Association (ELRA), S. 3172–3178.
- Zarriß, S. und J. Kuhn (2009). “Exploiting Translational Correspondences for Pattern-Independent MWE Identification”. In: *Proceedings of the Workshop on Multiword Expressions (MWE): Identification, Interpretation, Disambiguation and Applications*. Association for Computational Linguistics (ACL), S. 23–30.